

## Challenges to the Measurement of Estradiol: An Endocrine Society Position Statement

William Rosner, Susan E. Hankinson, Patrick M. Sluss, Hubert W. Vesper, and Margaret E. Wierman

Columbia University, College of Physicians and Surgeons (W.R.), and St Luke's-Roosevelt Hospital Center (W.R.), New York, New York 10019; Division of Biostatistics and Epidemiology (S.E.H.), School of Public Health and Health Science, University of Massachusetts Amherst, Amherst, Massachusetts 01003; Division of Network Medicine (S.E.H.), Brigham and Women's Hospital, Boston, Massachusetts 02115; Pathology Department (P.M.S.), Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts 02114; Clinical Chemistry Branch (H.W.V.), Division of Laboratory Sciences, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia 30341; University of Colorado School of Medicine (M.E.W.), Aurora, Colorado 80045; and Denver Veterans Affairs Medical Center (M.E.W.), Denver, Colorado 80220

**Objective:** The objective of the study was to evaluate the current state of clinical assays for estradiol in the context of their applications.

**Participants:** The participants were appointed by the Council of The Endocrine Society and charged with attaining the objective using published data and expert opinion.

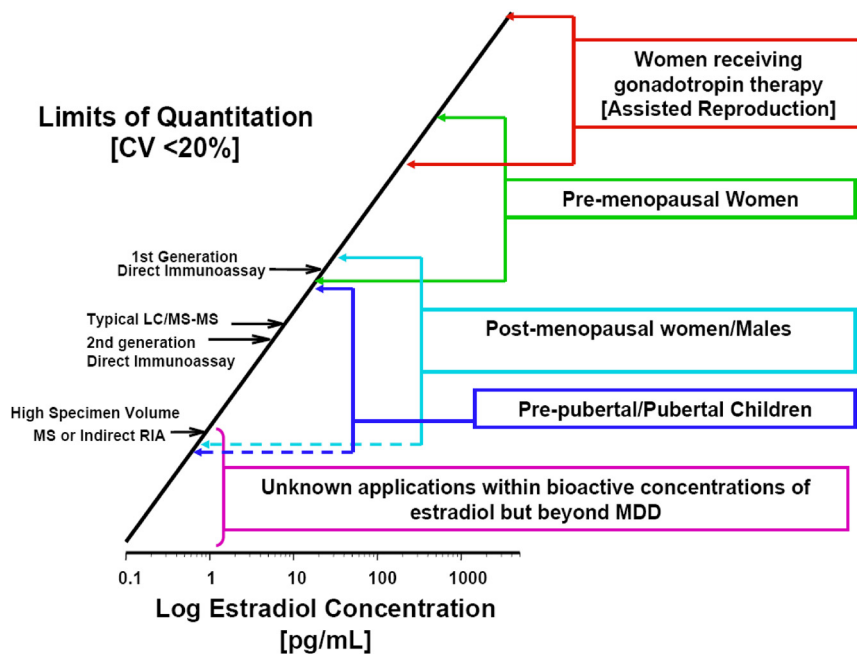
**Evidence:** Data were gathered from published sources via online databases (principally PubMed, Ovid MEDLINE, Google Scholar), and the clinical and laboratory experience of the participants.

**Consensus Process:** The statement was an effort of the committee and was reviewed by each member. The Clinical Affairs Committee, the Council of The Endocrine Society, and *JCEM* reviewers reviewed the manuscript and made recommendations.

**Conclusions:** The measurement of estradiol in biological fluids is important in human biology from cradle to grave. In addition to its centrality in sexual development, it has significant effects on skin, blood vessels, bone, muscle, coagulation, hepatic cells, adipose tissue, the kidney, the gastrointestinal tract, brain, lung, and pancreas. Alterations in its plasma concentration have been implicated in coronary artery disease, stroke, and breast cancer. Although modern immunoassays and liquid chromatography/tandem mass spectrometry-based methods for estradiol are reasonably well suited to the diagnosis and management of infertility (nonetheless, imprecision and method-to-method differences remain problematic), the very low concentrations that appear to be crucial in nonreproductive tissues are a separate and more difficult issue. Such levels of estradiol are too low to be routinely measured accurately or precisely, and further evolution of analytical methods and the way in which estradiol is standardized is needed. (*J Clin Endocrinol Metab* 98: 1376–1387, 2013)

Our understanding of the biology of sex hormones has undergone a remarkable change from the initial supposition that they were responsible only for the development and maturation of secondary sexual characteristics. It is difficult to identify a tissue that is not directly

influenced by estrogens and/or androgens. Despite the widespread importance of sex steroid hormones to human biology, our ability to measure them properly has not kept pace with their increasing importance in clinical medicine and research. For T measurement, this problem has been



**Figure 1.** Historical and current technical challenges to E2 assays with regard to analytical sensitivity. The primary methods used for clinical measurements are shown along the bar over a scale of E2 levels encompassing the situations in which measuring E2 in peripheral circulation is useful. (Detailed procedures for determining the limit of quantitation (LOQ) are available from the Clinical and Laboratory Standards Institute [CLSI EP17, <http://www.abrf.org/index.cfm?method=list.getAttachment&disclaimerAck=1&msg=83913&att=688>]). In this figure, a “rule-of-thumb” definition for the limit of quantitation (or minimal detectable dose [MDD]) is used; ie, the lowest concentration that can be measured with a coefficient of variation (CV) of 20% (19). It is useful to keep in mind that the 95% CI for a measure at this level is the mean  $\pm$  2 SD of repeated measurements. In other words, if the limit of quantitation is 80 pg/mL, values between 48 and 112 pg/mL cannot be distinguished with 95% confidence. The LOQ is typically dependent, among other things, on the operational condition of the instrument and the quality of reagents used; it therefore often differs in routine clinical laboratories when compared to the conditions under which the assay is first characterized and evaluated.

recognized by a position statement by The Endocrine Society (1) and a number of follow-up publications aimed at correcting the difficulties that were identified (2, 3). In this communication, we address the current problems with the measurement of estrogens, primarily estradiol (E2).

Accuracy, specificity, reproducibility (precision), and standardization of the measurement of E2 are critical. Uniform reference ranges for all ages, both sexes, and the great variety of normal and pathological conditions that influence the concentration of E2 in plasma are needed for optimal medical decision-making. Endocrinologists, clinicians, and researchers alike are hardly alone in facing difficulties in obtaining the reliable data that greatly influence their ability to make informed decisions. A recent issue of *Science* devoted its cover and a special section to data replication and reproducibility (4). Although the issues addressed therein are scientific only, replication and reproducibility in the clinical setting are of equal, if not greater, importance. If the physician changes the laboratory to which he/she sends samples, or if the patient goes to a different, more convenient lab, or if the laboratory

discards an old instrument and starts using a new and different one, shouldn't the E2 level being measured on the same sample return the same result in each of these situations? Of equal importance is the need for subsequent samples to be traceable to the same standard as the initial one, no matter where or how the measurement is performed. More often than not, that is not the case.

In this communication, we review the current state of how E2 is measured in the clinical and research communities and outline the importance of knowing plasma E2 concentrations in individuals and populations in normal physiology and in pathophysiological states.

## How Estradiol Is Measured

### Overview

Significant technical issues are associated with all currently available methods for the measurement of estradiol. When considered across the entire spectrum of available technologies, these issues encompass all of the key characteristics of analytical

methods: specificity, sensitivity, precision, and accuracy. These characteristics are inter-related; for example, method inaccuracy may be due to incorrect calibration and/or a lack of specificity (Figure 1).

### Methods for E2

Historically, methods for the measurement of E2 have followed the outline set out in Table 1. Quantitative measurements have been obtained by bioassay, mass spec-

**Table 1.** Evolution of Methods for Measuring Estradiol

Time Span	Methods
1930–1950	Extraction→liquid chromatography→bioassay
1950–present	Extraction→derivatization→gas chromatography→MS
1960–1980	Extraction→chromatography→RIA
1980–1990	Extraction→HPLC→UV absorption (240 to 280 nm)
1980–present	Direct RIA
1990–present	Automated direct RIA
1994–present	In vitro bioassay (research only)
2000–present	Extraction→HPLC→tandem MS

trometry (MS), UV absorbance, and immunoassay; only immunoassay and mass spectrometry methods have had the attributes necessary for clinical applications. To date, applying any of these measurement methods to complex biological specimens (eg, serum, plasma, tissue extracts) has required the isolation of steroids, both from other components of the specimen and from other steroids, using extraction and chromatography before the final measurement, eg, immunoassay, mass spectroscopy.

The original immunological method for quantifying E2 in clinical serum specimens was the RIA described by Abraham in 1969 (5, 6). This method consisted of extraction of serum with an organic solvent followed by quantification by RIA (such an assay is often referred to as “conventional RIA” or “indirect RIA”). Over the years, new immunological methods were developed that measure E2 in serum without its prior isolation, eg, no extraction or chromatography. These so-called “direct RIA” assays profoundly increased throughput. It is important to note that whereas no extraction is involved, the accuracy of direct immunoassays relies on a buffer that ensures the release of albumin and SHBG-bound E2. The method evolved further when radioisotopes were replaced with other technologies capable of being associated with very small masses, such as chemiluminescent or enzymatic labeling (7), which in turn served as a basis for greater automation and throughput. Today, most methods used in patient care measure E2 directly, without prior isolation from serum, with enzyme-based immunoassays (“direct assays”). A 2011 College of American Pathologists survey on instruments for the automated immunoassay of E2 lists 28 different assay platforms from 8 different companies (8).

Although a “gold standard” method for the measurement of E2, isotope dilution/gas chromatography (GC) coupled with mass spectrometry, was published about 25 years ago (9), that method was too complex, and its throughput too low for routine clinical use. Recently, simpler analytical methods using liquid chromatography (LC) coupled with mass spectrometry, eg, HPLC coupled to tandem mass spectrometry (MS/MS), have been developed and are increasingly used for measuring E2 in patient samples. These new methods, unlike the original RIA method but like the original MS method, use extraction plus a chromatographic step to separate E2 from similar compounds in the serum sample before quantification by mass spectrometry (10, 11). Mass spectrometry is increasingly used because it has the potential to provide greater specificity and sensitivity than immunological methods, which is especially important at the low E2 concentrations commonly observed in children, men, postmenopausal women, and women receiving aromatase inhibitors for the

treatment of breast cancer. However, at this time limited information is available about the discrepancies among individual mass spectrometry-based assays. Although these technologies have a number of theoretical advantages over immunoassays, they too may show considerable variability, as has been shown for other analytes such as T (12, 13). Furthermore, as is true for all methods, measurements based on mass spectrometry depend upon the accuracy of calibration, freedom from interference, and appropriateness of any analytical corrections (“recovery,” for example). Thus, it is important to realize that it is not solely the method per se that determines validity and utility.

Measurement of E2 in the clinical setting poses a number of difficulties for analytical methods. The assays need to be sensitive, specific, accurate, and precise over a wide concentration range. They need to:

- Measure E2 at very low concentrations. To monitor patients with breast cancer treated with aromatase inhibitors (to suppress endogenous estradiol levels), methods need to be able to distinguish between suppressed levels of less than 1 pg/mL and pretreatment levels that are commonly 10–15 pg/mL (7).
- Have a measurement precision that is sufficient to characterize a patient’s status and identify her response to treatment. For women taking aromatase inhibitors, methods will likely need to be sufficiently precise to distinguish between levels of less than 1 and 5 pg/mL.
- Measure reliably over a wide concentration range. E2 levels in elderly men and women are often < 5 pg/mL, whereas E2 testing performed in support of in vitro fertilization programs, for monitoring ovulation induction and ovarian hyperstimulation, require measurements to be reliable at levels of approximately 3000 pg/mL (14).
- Have a high specificity for E2. Not only is E2 converted to more than 100 conjugated and unconjugated metabolites (15–20), but also, patients may have circulating estrogens derived from exogenous sources, eg, conjugated equine estrogens, nutritional supplements, etc. Some of these compounds may cross-react with the antibody in the immunoassay or interfere with the HPLC/MS/MS measurements. Some of these compounds, such as estrone sulfate, occur in such relatively high concentrations in the circulation that even small cross-reactivities can result in profoundly deranged results. Other factors that may interfere with measurements are hemolyzed and lipemic samples (7) and the failure to release all protein-bound E2 before the final analysis.
- Provide accurate results that are comparable among different laboratories so that the data can be pooled to formulate clinical guidelines and allow results from in-

dividual patients to be compared with these guidelines. Similarly, accurate, reproducible data are essential for managing the patient whose testing is performed by several different laboratories using different methods.

### **Accurate, reproducible measurements across assays, time, and laboratories require the use of common reference materials**

A number of studies suggest that current routine clinical assays do not meet the needs set out above. Indeed, concerns about the analytical performance in the measurement of E2 among different assays and concentrations have been reported for over 20 years (21–23).

The limit of quantitation of most direct immunoassays ranges from 30 to 100 pg/mL, which is insufficient for measurements in children, postmenopausal women, men, and women taking aromatase inhibitors (7). Even mass spectrometry-based methods and indirect RIAs that have been recommended for such low concentrations may have difficulty with levels < 5 pg/mL (7, 24, 25).

Concerns about the specificity of current E2 assays have been raised for many years (10, 11). Findings from 1 study indicate that interfering compounds may cause E2 measured values to be 10 times higher than the true value (7). This study also suggests that the concentration of interfering substances in patient samples can vary greatly from patient to patient. For the same reasons, it was suggested that E2 plasma levels, using some direct immunoassays, measured in women taking aromatase inhibitors are most likely artifacts rather than true E2 values (25). This is of particular importance because the failure or success of the treatment may wrongly be attributed to a serum E2 concentration that does not reflect the true effect of the aromatase inhibitor.

Another study, using data from 159 laboratories enrolled in the New York State Department of Health proficiency testing program, reported substantial interferences from conjugated estriol. These interferences resulted in high variability and inaccuracy in E2 measurements. A number of assays also had substantial interferences from unconjugated estriol, leading to positive and/or negative biases of up to 60%. The authors found inconsistencies in the use of manufacturer-recommended dilution protocols that would have minimized some of the observed variability (26).

The precision of E2 measurement has been reported to be insufficient, especially at low concentrations. The performance of 7 automated assays over 14 months showed that the method-specific variability at low E2 concentration ranged from 7.5 to 28.4% coefficient of variation. The authors concluded that with most assays, very low E2 concentrations, as observed in postmenopausal women,

can be determined only with a precision inadequate for clinical assessments (27). Similarly, data from a UK external quality assessment program from 103 participants, using mainly commercial direct assays, showed high variability within laboratories using the same assay (28). Another study comparing 8 different immunoassays to a GC/MS method found inconsistent assay precision and accuracy. Although some assays showed imprecision of less than 10% at E2 levels commonly observed in postmenopausal women and men (18 pg/mL), as well as levels observed during ovarian stimulation (200–2000 pg/mL), others performed inadequately. Although some assays showed good agreement with the GC/MS method, others showed inaccuracy that might improve with additional method optimization and calibration standardization (29).

In 2007, 140 laboratories measured E2 in serum samples as part of a Belgian external quality assessment scheme. They compared immunoassay results against a GC/MS-based method and found coefficients of variation ranging from 4 to 49% and bias (disagreement with GC/MS) ranging from 26 to 239% for individual samples measured with the same assay. No single method had bias values of less than 10% for all samples. Although bias and imprecision increased with decreasing E2 concentration, other assays showed differences in performance that appeared independent of the sample E2 concentration. In fact, the highest bias of 239% was observed in a sample with a target concentration of 553 pg/mL, indicating interferences by other compounds (30).

Some improvements in the performance of immunoassays have been made, as indicated in more recent comparison studies (29, 31). These changes, however, are not sufficient to meet current needs in research and patient care. To overcome these problems and ensure better patient care, standardization of E2 measurements has been suggested (25, 29, 32–34). Accurate results across methods are essential for the formulation of evidence-based practice guidelines and the management of patients receiving care at multiple locations, supported by different laboratories, using different methods.

The significance of how blood samples are processed and the nature of standards used in estradiol assays is not generally appreciated in the clinical community. Precise and accurate weighing of crystalline E2 is simple. When that material is dissolved in a liquid, whether it be plasma or a buffer, and then extracted (or measured directly) and measured by immunoassay or mass spectrometry, the results will depend on how the assay has been standardized and the conditions to which the sample has been subjected. A variety of influences including, but not limited to, the nature of the tube in which the specimen is drawn and



obtained, the mode of extraction, the length of storage, etc, may influence the result. Standards for assay calibration are supplied in artificial media that hopefully behave like plasma, but may not. More studies are needed using fresh-frozen, unaltered sera to better assess the sources of measurement variability. Indeed, preanalytical issues and errors represent a significant source of inaccuracy (35). The importance of the preceding issues cannot be overstated. Indeed, it is clear that we need a single standard for E2 to which all measurements can be traced, and that standard should be available, not in an artificial medium, but in serum or plasma or whatever specimen type is being measured, eg, urine, tissue extract, saliva, etc. Agreement with the standard will ensure that the myriad of technical difficulties has been overcome. The way to arrive at this standardization has been achieved for T and is described at the web site of the Centers for Disease Control and Prevention ([http://www.cdc.gov/labstandards/hs\\_standardization.html](http://www.cdc.gov/labstandards/hs_standardization.html)).

The salient features of difficulties and solutions to estradiol measurements can be briefly summarized:

- Because of the wide variety of physiological circumstances and disease states of interest, assays for estradiol must be capable of yielding results over a very large range of concentrations, eg, a factor 10 000.
- Historically, sensitive assays for estradiol have been immunologically based. The goodness of these assays is related to many factors, but most have some difficulty with specificity and insufficient sensitivity to satisfy all clinical and scientific requirements.
- Most recently, methods based on mass spectroscopy appear to be suitable for addressing some of the shortcomings of immunoassays.
- We emphasize that whatever method is used, we need a single standard for E2 to which all measurements can be traced, and that standard should be widely available.

## The Measurement of Estradiol in Clinical Practice and Translational Research

Plasma E2 levels vary widely, depending on a host of normal and pathological conditions, each of which is important in its own context. Here, we review the complexity of that variety to establish the importance of this measurement across a wide swath of human biology.

### Children

Sensitive and specific assays for E2 are needed to track the onset of pubertal development and better define the ranges of normal levels across childhood. Current direct E2 assays cannot accurately measure with reasonable pre-

cision E2 below 30 pg/mL; thus, most E2 levels are undetectable until Tanner stage 2–3 of puberty (36). Because of this lack of sensitivity, the role of estrogens in prepubertal growth and their impact in boys compared to girls remains largely unexplored.

Ultrasensitive E2 assays, perhaps suitable for use in children, were developed in the 1990s but were time-consuming, required large amounts of serum, and at low levels were subject to the same lack of specificity as standard RIA (37, 38). Klein et al (39, 40) developed and refined a cell-based bioassay with a sensitivity of 0.2 pg/mL to monitor low levels of estrogenic activity, but this method does not measure a distinct chemical entity, requires a pre-extraction, and is dependent on standardization of temperature, cell density, and incubation parameters, making applications in clinical practice impractical (41). Recently, Shi et al (42) used isotope dilution/GC/MS of 24-hour urinary samples in a longitudinal assessment of specific estrogens 1 and 2 years before the onset of the growth spurt. Girls were assessed yearly from age 5 and boys from age 6. Of 376 subjects, data were available on 120. The sum of estrone, E2, and estriol was used as a marker of estrogen production. In contrast to their predictions, urinary estrogens and/or their metabolites did not predict initiation of the pubertal growth spurt, but rather predicted the duration of pubertal growth (42). Higher levels of estrogens at earlier ages correlated with the onset of breast development in girls but had no apparent role in pubertal maturation in boys. Importantly, however, a substantial number of samples were below the limit of detection of these new urinary estrogen assays. To date, no study of serum levels of E2 or other estrogens across childhood, using methods with adequate sensitivity, is available. Thus, many issues related to the role of estrogens in pubertal development and growth remain unanswered. In addition to understanding normal puberty, the understanding of E2 levels in precocious and delayed puberty would be served by improved E2 assays. Similarly, the response to and efficacy of GnRH agonists for the treatment of precocious puberty and, conversely, therapies to induce pubertal development require precise and sensitive E2 measurement.

Rarely, gonadal tumors that secrete human chorionic gonadotropin, T, and/or E2 produce clinical signs of estrogen excess, but with plasma E2 levels that are often below the detection limit of the assay (43). Boys who present with pubertal gynecomastia, which implies an imbalance of T and E2, cannot be distinguished by current hormonal assays (44).

Thus, current E2 assays are inadequate to define the normal stages of puberty and assess the effects of estrogens on growth in the prepubertal child, and they are of limited

use in pathological disorders of the pubertal process. The ability to measure basal E2 levels in these disorders and changes in E2, consequent to therapeutic interventions, would aid clinicians and researchers alike.

## Adult women

### **Normal menstrual cycle and amenorrhea**

Estradiol levels, in conjunction with measurements of FSH and LH, define the stage of the menstrual cycle; day 1 of the menstrual cycle is associated with an E2 of greater than 60 pg/mL (45, 46) as ascertained by RIA after extraction and chromatography. Persistently lower levels are associated with hypoestrogenic effects at target tissues such as loss of bone mass (47). Unfortunately, current direct E2 assays, used to measure levels in normal cycling women in the follicular phase, are insensitive below 20 pg/mL, making the diagnosis and treatment of E2 deficiency quite difficult (47).

In contrast, E2 levels in the early follicular phase (d 1–5 of the menstrual cycle), using a mass spectroscopy-based assay, were  $68.1 \pm 18.6$  pg/mL (48). These data suggest that newer mass spectrometry-based assays demonstrate E2 levels across the menstrual cycle similar to those found using RIAs that used extraction and chromatography, and higher than the direct E2 values obtained from immunoassays based on commercial instruments used in clinical laboratories. The ability to define normal ranges in women across the menstrual cycle is critical for defining estrogen deficiency or excess in premenopausal women.

### **Reproductive technology**

Sensitive and specific E2 measurements are needed for women undergoing induction of ovulation. In this case, assays sufficiently sensitive to measure low baseline E2 levels as well as high levels (ie, 250–2000 pg/mL) are needed to assess the efficacy of treatment, the timing of human chorionic gonadotropin administration to trigger ovulation, and cutoffs designed to abort cycles that risk hyperstimulation. A comparison of E2 levels, using different commercial immunoassay kits, in 41 women undergoing induction of ovulation showed that the most sensitive assay (55 pM; 15 pg/mL) also had a higher precision than other assays (49). However, for higher E2 levels, serial dilution was often required, limiting the goal of a fast processing time for samples containing the high E2 levels monitored after ovarian stimulation. Thus, robust E2 assays across several orders of concentration are needed for use in reproductive technology.

### **Pregnancy**

Monitoring of E2 levels across pregnancy can be of use in pregnancy-related disorders. E2 levels together with

progesterone and inhibin B levels predict the risk of hydatidiform mole in patients with molar pregnancies (50). Levels of E2 correlated with disease activity in pregnant women with systemic lupus erythematosus (51). Current direct E2 assays do not provide this discrimination.

### **Menopausal hormone therapy**

Measurement of an E2 level, together with FSH and/or anti-Müllerian hormone, would be useful to predict the timing of the last menstrual cycle. The Study of Women's Health Across the Nation (SWAN) collected serum annually for the measurement of E2 and FSH (52). Mean E2 levels did not decrease until 2.03 years before the final menstrual period, using a sensitive immunoassay with a lower limit of detection of 1 pg/mL (53). In contrast, most commercial E2 assays cannot accurately measure E2 levels in postmenopausal women (<20 pg/mL).

In menopausal women using hormonal therapy, adjustment of the dose of E2 to the lowest level that controls symptoms has been recommended. The ability to accurately measure E2 levels in women given oral or transdermal E2 would help to determine compliance, efficacy, and achievement of specific levels and would be of great assistance to clinicians. Such efforts are under way. Comparison of an extraction-based (indirect) (53) and a non-extraction-based (direct) immunoassay with a mass spectrometry-based assay in 40 postmenopausal women showed interesting similarities and differences. Three different indirect assays correlated better with GC/MS/MS than a direct assay. In a larger cohort (n = 374), indirect and direct RIAs overestimated E2 levels by 14 and 68%, respectively, and were less reproducible than GC/MS/MS (53). A GC/MS/MS-based assay compared E2 levels in premenopausal women in the early follicular phase ( $55.4 \pm 10.3$  pg/mL) to postmenopausal women less than ( $4.9 \pm 1.3$  pg/mL) and more than ( $1.3 \pm 0.3$  pg/mL) 5 years after the last menstrual period. They demonstrated a continuing decline in serum E2 with age as well as the expected decrease with the menopausal transition (48). With these newer approaches, investigators and clinicians can determine the optimal dose and threshold of response of various target tissues to estrogen therapy, eg, bone, lipids, cardiovascular risk markers, cognition, and mood.

### **Hormone-dependent malignancies and ablation/replacement models**

Women with endometriosis or leiomyoma who receive GnRH agonists to induce medical castration, together with a low-dose estrogen add-back regimen, are disadvantaged by the absence of a more sensitive and specific E2 assay to aid in achieving a specific plasma E2 target. Similarly, women treated with aromatase inhibitors (to de-

crease circulating estrogens) for breast cancer need such assays to better define the optimal level of suppression, both to control the disease state and to define a replacement strategy that would optimize the target organ effects of E2 deficiency. In women taking aromatase inhibitors, most RIA methods for E2 detect cross-reacting estrogen metabolites resulting in higher measured E2 levels than those measured by GC/MS-based assays; at this time, a GC/MS method should be used to monitor therapy in women with breast cancer (54).

## Adult men

### **Normal physiology and pathophysiology**

The role of target tissue aromatization of T to E2 in men is poorly understood. Estradiol levels in men are lower than in premenopausal women (53); they are at or below the sensitivity of available direct E2 assays (55). In addition, there is concern about the lack of specificity at the lower ranges of the assay, similar to issues in the measurement of levels in postmenopausal women. Comparison of the measurement of 12 steroids by GC/MS or LC/MS/MS and 6 steroids by RIA in fasting samples from 20 healthy men (aged 50–65 y) revealed significant differences. RIA showed higher E2 levels, suggesting greater cross-reactivity compared to other methods ( $33 \pm 6$  compared to  $21 \pm 5$  pg/mL). Although each assay showed good reproducibility, the authors caution about comparing absolute levels across assays or studies (55). Thus, concordance in assays is required to define normal and abnormal E2 levels in males.

### **Bone**

Data suggest that E2 levels are better predictors of bone mass in men than T, but only when ultrasensitive E2 assays, not commercially available, are used (56). In another study, E2 levels measured by a mass spectroscopy-based method were compared to those assessed by a platform-based immunoassay in 313 men in relation to volumetric bone mineral density at various skeletal sites (57). There was not a good correlation between the two types of assays ( $r = 0.63$ ), although both correlated equally well with volumetric bone mineral density. Thus, sensitive E2 measurements may be used in the future to help predict the risk of osteoporosis and fracture in men.

### **Muscle strength**

Newer sex hormone assays have been used recently to dissect the role of T and E2 in men in relation to strength and performance. Sex hormone levels were measured in 1489 men > 64 years old in relation to muscle mass, hand-grip strength, gait speed, step length, and chair-stand test as measures of physical performance (58). E2 levels cor-

related with muscle mass, as seen in prior studies (59), but were inversely correlated with muscle strength. More sensitive and specific E2 assays will expand our understanding of estrogen action at different target tissues in men.

### **Hypogonadism**

Estradiol levels in 91 men with idiopathic hypogonadotropic hypogonadism (IHH) and Klinefelter's syndrome were compared to normal controls using a direct RIA with a stated detection limit of 2 pg/mL (7.3 pM). E2 levels in IHH ranged from 4–68 pg/mL (median, 12 pg/mL) (60). In 24 of the subjects, the direct RIA levels were compared with a GC/MS/MS-based method; the correlation was 0.969, suggesting that some direct assays may compare favorably with mass spectrometry-based ones. However, almost 50% of the samples from men with IHH were below the detectability of this assay (no data were provided on the performance of the GC/MS/MS assay comparison). Sensitive specific E2 assays would help to define the role of estrogens in men with different types of hypogonadism.

### **Hormone-dependent malignancies and response to hormonal manipulation**

Androgens are obligate precursors of estrogens. Thus, in men with prostate cancer undergoing androgen deprivation therapy, E2 measurements may be useful predictors of the effects of this therapy on targets such as bone, heart, and metabolic status. Indeed, it appears that estrogens are involved in both prostate cancer and benign prostatic hyperplasia (61); careful measurement of E2 in men with and without androgen deprivation should aid our understanding of its role in disease and predicting risk of the complications of sex hormone deficiency after androgen deprivation.

In summary, the care of patients across the life span is hampered by the lack of availability of sensitive, precise, and specific E2 assays. Thus, clinicians are receiving potentially spurious data derived from direct E2 assays performed on platforms in hospital or clinical laboratories. Mass spectrometry-based methods are arguably stronger technically and show better performance characteristics than currently available direct immunoassays, although agreement among mass spectrometry-based methods is often less than perfect. In any event, our criteria for assay acceptability must not be based on the method used, whether it be immunoassay-based or mass spectrometry-based, but rather on agreed-upon standardization, so that clinically meaningful concentrations can be measured accurately and reproducibly across laboratories and platforms. Definition of the boundaries of normal physiology, definition of pathological states, and optimal targeting of

therapeutic replacement strategies would be served better by improved E2 assays.

We should not neglect to mention that at this time the feasibility/usefulness of an assay is also dependent on cost, sample size, and assay availability. These are real problems confronting many investigators, and they need to be addressed.

## The Measurement of Estradiol in Epidemiological Research

The use of E2 assays in epidemiological research has a number of the same purposes as in the clinical/translational research described above. Clinical data are used to predict what might befall the individual patient; epidemiological data are the sum of what occurs on average in the many individuals that comprise a population. The primary goals of large epidemiological data sets are to: describe the distribution of E2 levels in varying populations, eg, children (62), women (63, 64), and men (65); evaluate lifestyle, environmental, and genetic predictors of these levels; and assess how E2 levels influence both disease risk and patient survival. In addition, in the long term, such data can improve our understanding of the contribution of E2 levels to disease etiology and help to determine whether these levels can be used clinically to stratify patients as high or low risk for a given disease.

Breast cancer is an excellent example of the use of E2 measurements in epidemiological studies. E2 levels play a primary role in breast carcinogenesis (66). A number of prospective epidemiological studies have shown that increased circulating E2 levels in postmenopausal women are significantly associated with subsequent breast cancer risk (67–69). Relative risks (RRs) range from 1.5–3.0, comparing women with the highest vs the lowest E2 levels. Associations of E2 and breast cancer in premenopausal women remain uncertain and have been difficult to dissect because of both the variation in E2 across the menstrual cycle and limitations of the assays (70). Other studies have been conducted to evaluate differences in hormones in women by ethnicity (64) to determine whether differences in endogenous hormone levels could be related to variations in breast cancer rates among women of various ethnicities. Furthermore, a number of investigators have assessed the correlation between plasma hormone levels and alcohol intake (71) or conducted randomized feeding studies of alcohol to examine changes in hormone levels (72) to help determine the underlying mechanism(s) that have linked alcohol intake to increased breast cancer risk in women (73). Finally, recent studies have addressed the correlation between plasma E2 levels and multiple poly-

morphisms in hormone-metabolizing genes to better understand genetic contributions to circulating endogenous E2 levels (74).

Similar to the work in breast cancer, a wide range of epidemiological studies have been conducted to assess the associations between circulating E2 levels and the risk of other chronic diseases including endometrial cancer (75, 76), cardiovascular disease (77, 78), cognitive dysfunction (79), and fracture (80).

In addition to examining circulating E2 levels, epidemiological studies have utilized several tissues (eg, breast aspirate fluid [81] or urine [71]) where similar measurement issues exist.

Methodological concerns in E2 assays raise important basic issues in epidemiology. The use of highly sensitive, specific, and precise E2 assays in epidemiological studies is critical for several reasons. First, most studies evaluate differences across the continuum of normal levels rather than discerning normal from abnormal levels, as is frequently done clinically. For example, in a pooled analysis of 9 combined prospective studies on breast cancer, the median difference in E2 levels between cases and controls was only 6.3% (range, –9.2 to 33.6%), yet these differences translated into a 2-fold higher RR of breast cancer when comparing the top vs bottom 20% of E2 levels (RR = 2.0; 95% confidence interval [CI], 1.5–2.7) (67). Hence, relatively modest differences in levels between the groups being compared are often of interest and can translate into important predictors of disease risk.

Imprecision in E2 assays can result in attenuated/absent measures of association, suggesting a weaker association or even no association with disease (or other outcome) than truly exists. For example, in a study where the precision of E2 was assessed in 4 different laboratories using varying assays, the coefficients of variation ranged from 8–59% (82). The calculated potential impact of this range of laboratory error on the RR was substantial, ie, true RR of 2.5 could be attenuated to 1.1 to 1.6 (the actual RR observed in the study). In addition, for both logistic and financial reasons, in most epidemiological studies only a single blood sample can be collected per study subject, although in studies of chronic disease, long-term average hormone levels are generally of greatest interest. When only 1 blood sample can be used to characterize these levels, it is critical that both assay precision and preanalytical variables be optimized. Several prior studies, described in greater detail previously in this section, have documented the variable precision of E2 assays.

We are unaware of examples using different E2 assays within the same epidemiological study when evaluating an association between E2 and disease risk. However, in a pooled analysis of 9 prospective studies of E2 and breast



cancer risk (67), stratified by assay type, the RR associated with a doubling of E2 levels from the 4 studies using various direct assays was somewhat lower than the RR from studies using assays with a prior extraction step. However, this difference was not statistically significant (direct assays, RR = 1.23 [95% CI, 1.04–1.44]; vs assays with a prior extraction step, RR = 1.35 [95% CI, 1.15–1.58]). Interestingly, this same consortium recently reported (pooling data from 13 studies) on the correlation between postmenopausal E2 levels and body mass index (BMI), a variable known to be strongly correlated with E2 levels (83). Significant heterogeneity in the association was noted among studies, and some of this was due to the assay used. E2 was 82% higher in obese vs lean women in studies that used assays with a prior purification step and only 30% higher in studies that used a direct, nonextraction assay. Similar results had been previously reported (53), where correlations between BMI and E2 ranged from approximately 0.65 (for GC/MS/MS and an extraction-based RIA assay) to 0.25 (for a direct immunoassay).

The varying sensitivity and specificity of different laboratory assays also makes comparisons of results across epidemiological studies difficult/unreliable. For example, in studies of plasma E2 in postmenopausal women and breast cancer risk, median levels in control subjects ranged from 5.9 to 27.5 pg/mL (67), over a 4-fold range. Similarly, in studies of E2 and prostate cancer in men, median levels in control subjects ranged from 18 to 60 pg/mL, a 3-fold range (84). It is highly unlikely that differences of this magnitude between research studies are due to varying attributes of the study population (eg, ethnicity, different average BMI, etc); hence, some, if not most, of these differences in E2 levels are likely due to assay differences. In this instance, assay methodologies varied substantially across studies (eg, both direct immunoassays and indirect immunoassays that included a prior extraction and separation step were used). Inadequate sensitivity also can limit the interpretation of epidemiological studies of E2 and disease risk. For example, in a prospective evaluation of endogenous E2 levels in postmenopausal women and vertebral fractures (85), 33% of the levels were < 5 pg/mL, the assay limit of detection (RIA after extraction and HPLC separation); actual E2 concentrations on the 33% of the population with values below the sensitivity of the assay would likely have added additional insight into the strength of the association and shape of the dose-response curve. An additional longer-term goal of epidemiological research on hormones and disease is to determine whether hormone levels could be used to help identify persons at high risk of disease who might benefit from risk reduction strategies (similar to what is routinely done with cholesterol measurements). Again, an inability to compare as-

sociations or absolute levels between studies hampers this goal.

Thus, although we have learned much regarding the association between E2 and a range of health outcomes, the different assays used and variable assay quality have clearly limited the conclusions that could be drawn from epidemiological studies. Improved E2 assays that are more sensitive, accurate, and precise would allow proper comparisons across studies and would greatly enhance these efforts.

In summary, we cannot state too strongly the need for accurate, specific, reference-based standards in epidemiological studies. Such studies contribute greatly to decisions for counseling about and treating a panoply of illnesses in individual patients. Breast cancer, diseases of bone, cognitive dysfunction, and cardiovascular disease are among those that suffer from a limited ability to combine data from diverse studies because measurements and standards are not uniform. Furthermore, clinical guidelines become meaningless in the face of assays in individuals or populations that are different from those used in developing the guidelines.

## Conclusion

The measurement of estradiol concentrations in plasma (as well as other biological fluids) has assumed an importance in human biology from cradle to grave. Its import on sexual development and secondary sexual characteristics is what led to its discovery and its grouping as a “sex hormone.” It is now known to have important effects on skin, blood vessels, bone, muscle, coagulation, hepatic cells, adipose tissue, the kidney, the gastrointestinal tract, brain, lung, and pancreas. Alterations in its plasma concentration have been implicated in coronary artery disease, stroke, and breast cancer, all among the more prevalent causes of mortality in people. Effects of E2 in these tissues are likely at peripheral concentrations below the current detection limits of essentially all methods (2 to 20 pg/mL); indeed, levels below 0.2 pg/mL are likely to be meaningful considering the affinity constant of estradiol receptors (3 to 40 pM in breast cancer) (86, 87).

Currently, clinical E2 measurements are an essential component of the diagnosis and management of infertility. Modern immunoassays and LC/MS/MS-based methods are well suited to these applications, as well as to the less frequent but critical identification of E2-secreting tumors (where relatively high concentrations are being measured), although imprecision and method-to-method differences remain problematic. However, the very low concentrations that appear to have import in nonreproductive

tissues are a separate and more difficult issue. Such levels of E2 are too low to be routinely measured accurately or precisely, and further evolution of analytical methods and the way in which E2 is standardized is needed.

What is needed is more easily stated than accomplished, but we recommend:

- A universally recognized E2 standard to which all measurements can be traced.
- Reference ranges for E2 that are age- and gender-specific.
- Reference ranges for E2 that are biologically specific:
  1. Stage of puberty/adolescence
  2. Stage of menstrual cycle
  3. Stage of menopause
- A wider recognition among physicians, laboratorians, and investigators of the fact that the low E2 values seen in men, children, and menopausal women, which are obtained from direct assays, eg, those done on whole serum without prior extraction, are not likely to be trustworthy.
- New methods capable of accurately and precisely measuring estradiol levels in the 0.2 to 2 pg/mL range in routine clinical specimens. Until such methods are available, a system needs to be in place that allows the continuous evaluation of existing methods and facilitates the improvement of these methods.

### Barriers to recommendations

The following barriers are substantive but each can be overcome. This is now being done as an outgrowth of The Endocrine Society's position statement on T (1). The division into categories based on finances is somewhat artificial because all the "nonfinancial" barriers have a dollar cost.

### Financial

- Establishing a universal accuracy-based standard will involve new spending.
- Changing assays to more expensive ones, eg, from immunoassay to mass spectroscopy will increase costs.
- Developing superior immunoassay-based methods will involve increased outlays for research to accomplish this end.

### Political/educational/inertial/scientific

- Educating physicians to insist on accuracy-based measurements will involve a substantive educational effort. This will be costly and will have to overcome historical habits and attitudes.

- Convincing journals to insist on accuracy-based measurements for all publications. This represents yet another chore for journals and will require each to set a deadline date beyond which inappropriate methods will no longer be acceptable.
- Convincing government and third-party payers that a more expensive assay is more cost effective than a less expensive one that provides incorrect information.
- Establishing all the reference ranges that we suggest will be both time-consuming and expensive.

### Acknowledgments

Address all correspondence and requests for reprints to: William Rosner, MD, St. Luke's/Roosevelt Hospital Center, Department of Medicine, Endocrinology, AJA 403, 1000 10th Avenue, New York, New York 10019. E-mail: wr7@columbia.edu.

Disclosure Summary: S.E.H. is funded by National Institutes of Health Grants R01 CA49449 and R01 CA6762. W.R., P.M.S., H.W.V., and M.E.W. have no conflicts to declare.

### References

1. Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H. Position statement: utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. *J Clin Endocrinol Metab.* 2007;92:405–413.
2. CDC workshop report improving steroid hormone measurements in patient care and research translation. *Steroids.* 2008;73:1285–1352.
3. Rosner W, Vesper H. Preface. CDC workshop report improving steroid hormone measurements in patient care and research translation. *Steroids.* 2008;73:1285.
4. Data replication and reproducibility. *Science.* 2011;334:1225–1233.
5. Abraham GE, Odell WD. Solid-phase immunoassay for estradiol 17- $\beta$ : a semi-automated approach. In: Peron FG, Caldwell BV, eds. *Immunologic methods in steroid determination.* New York: Appleton-Century-Crofts; 1970;87–112.
6. Abraham GE. Solid-phase immunoassay for estradiol 17- $\beta$ . *J Clin Endocrinol Metab.* 1969;29:866–870.
7. Stanczyk FZ, Jurow J, Hsing AW. Limitations of direct immunoassays for measuring circulating estradiol levels in postmenopausal women and men in epidemiologic studies. *Cancer Epidemiol Biomarkers Prev.* 2010;19:903–906.
8. *Laboratory instrumentation product guide: immunoassay analyzers.* Northfield, Illinois: College of American Pathologists; 2011; 17–54.
9. Thienpont LM, Verhaeghe PG, Van Brussel KA, De Leenheer AP. Estradiol-17  $\beta$  quantified in serum by isotope dilution-gas chromatography-mass spectrometry: reversed-phase C18 high-performance liquid chromatography compared with immuno-affinity chromatography for sample pretreatment. *Clin Chem.* 1988;34:2066–2069.
10. Stanczyk FZ, Clarke NJ. Advantages and challenges of mass spectrometry assays for steroid hormones. *J Steroid Biochem Mol Biol.* 2010;121:491–495.
11. Giese RW. Measurement of endogenous estrogens: analytical challenges and recent advances. *J Chromatogr A.* 2003;1000:401–412.
12. Thienpont LM, Van UK, Blincko S, et al. State-of-the-art of serum

- testosterone measurement by isotope dilution-liquid chromatography-tandem mass spectrometry. *Clin Chem*. 2008;54:1290–1297.
13. Vesper HW, Bhasin S, Wang C, et al. Interlaboratory comparison study of serum total testosterone [corrected] measurements performed by mass spectrometry methods. *Steroids*. 2009;74:498–503.
  14. Demers LM. Testosterone and estradiol assays: current and future trends. *Steroids*. 2008;73:1333–1338.
  15. Lepage J, Bernard O, Plante M, et al. Specificity and regioselectivity of the conjugation of estradiol, estrone, and their catecholestrogen and methoxyestrogen metabolites by human uridine diphosphoglucuronosyltransferases expressed in endometrium. *J Clin Endocrinol Metab*. 2004;89:5222–5232.
  16. Dawling S, Roodi N, Mernaugh RL, Wang X, Parl FF. Catechol-O-methyltransferase (COMT)-mediated metabolism of catechol estrogens. *Cancer Res*. 2001;61:6716–6722.
  17. Lippert TH, Seeger H, Mueck AO. The impact of endogenous estradiol metabolites on carcinogenesis. *Steroids*. 2000;65:357–369.
  18. Lee AJ, Cai MX, Thomas PE, Conney AH, Zhu BT. Characterization of the oxidative metabolites of 17 $\beta$ -estradiol and estrone formed by 15 selectively expressed human cytochrome p450 isoforms. *Endocrinology*. 2003;144:3382–3398.
  19. Ekins R, Edwards P. On the meaning of “sensitivity”. *Clin Chem*. 1997;43:1824–1831.
  20. Mueck AO, Seeger H, Lippert TH. Estradiol metabolism and malignant disease. *Maturitas*. 2002;43:1–10.
  21. Schioler V, Thode J. Six direct radioimmunoassays of estradiol evaluated. *Clin Chem*. 1988;34:949–952.
  22. Nisbet JA, Jomain PA. Discrepancies in plasma estradiol values obtained with commercial kits. *Clin Chem*. 1987;33:1672.
  23. Thomas CM, van den Berg RJ, Segers MF. Measurement of serum estradiol: comparison of three “direct” radioimmunoassays and effects of organic solvent extraction. *Clin Chem*. 1987;33:1946–1947.
  24. Bhasin S, Cunningham GR, Hayes FJ, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2010;95:2536–2559.
  25. Dowsett M, Folkard E. Deficits in plasma oestradiol measurement in studies and management of breast cancer. *Breast Cancer Res*. 2005;7:1–4.
  26. Cao Z, Swift TA, West CA, Rosano TG, Rej R. Immunoassay of estradiol: unanticipated suppression by unconjugated estriol. *Clin Chem*. 2004;50:160–165.
  27. Reinsberg J, Batz O, Bertsch T, et al. Precision and long-term stability of different estradiol immunoassays assessed in a multi-center quality control study. *Clin Lab*. 2009;55:201–206.
  28. Middle JG. External quality assessment of steroid hormone assays in the United Kingdom. *Ann Ist Super Sanita*. 1991;27:459–465.
  29. Yang DT, Owen WE, Ramsay CS, Xie H, Roberts WL. Performance characteristics of eight estradiol immunoassays. *Am J Clin Pathol*. 2004;122:332–337.
  30. Coucke W, Devleeschouwer N, Libeer JC, Schiettecatte J, Martin M, Smits J. Accuracy and reproducibility of automated estradiol-17 $\beta$  and progesterone assays using native serum samples: results obtained in the Belgian external assessment scheme. *Hum Reprod*. 2007;22:3204–3209.
  31. Sluss PM, Hayes FJ, Adams JM, et al. Mass spectrometric and physiological validation of a sensitive, automated, direct immunoassay for serum estradiol using the Architect. *Clin Chim Acta*. 2008;388:99–105.
  32. Thienpont LM, De Leenheer AP. Efforts by industry toward standardization of serum estradiol-17  $\beta$  measurements. *Clin Chem*. 1998;44:671–674.
  33. Thienpont L. Meeting report: First and Second Estradiol International Workshops. *Clin Chem*. 1996;42:1122–1124.
  34. Stanczyk FZ, Lee JS, Santen RJ. Standardization of steroid hormone assays: why, how, and when? *Cancer Epidemiol Biomarkers Prev*. 2007;16:1713–1719.
  35. Raff H, Sluss PM. Pre-analytical issues for testosterone and estradiol assays. *Steroids*. 2008;73:1297–1304.
  36. Bay K, Andersson AM, Skakkebaek NE. Estradiol levels in prepubertal boys and girls—analytical challenges. *Int J Androl*. 2004;27:266–273.
  37. Carlstrom K. Low endogenous estrogen levels—analytical problems and tissue sensitivity. *Acta Obstet Gynecol Scand Suppl*. 1996;163:11–15.
  38. Ikegami S, Moriwake T, Tanaka H, et al. An ultrasensitive assay revealed age-related changes in serum oestradiol at low concentrations in both sexes from infancy to puberty. *Clin Endocrinol (Oxf)*. 2001;55:789–795.
  39. Klein KO, Baron J, Colli MJ, McDonnell DP, Cutler GB Jr. 1994 Estrogen levels in childhood determined by an ultrasensitive recombinant cell bioassay. *J Clin Invest*. 94:2475–2480.
  40. Larmore KA, O'Connor D, Sherman TI, Funanage VL, Hassink SG, Klein KO. Leptin and estradiol as related to change in pubertal status and body weight. *Med Sci Monit*. 2002;8:CR206–CR210.
  41. Rosner W. La plus ça change, la plus c'est la même chose—will bioassays make a comeback? *J Clin Endocrinol Metab*. 2001;86:1898–1899.
  42. Shi L, Remer T, Buyken AE, Hartmann MF, Hoffmann P, Wudy SA. Prepubertal urinary estrogen excretion and its relationship with pubertal timing. *Am J Physiol Endocrinol Metab*. 2010;299:E990–E997.
  43. Fossa SD, Klepp O, Barth E, Aakvaag A, Kaalhus O. Endocrinological studies in patients with metastatic malignant testicular germ cell tumours. *Int J Androl*. 1980;3:487–501.
  44. Large DM, Anderson DC. Twenty-four hour profiles of circulating androgens and oestrogens in male puberty with and without gynecomastia. *Clin Endocrinol (Oxf)*. 1979;11:505–521.
  45. Soules MR, McLachlan RI, Ek M, Dahl KD, Cohen NL, Bremner WJ. Luteal phase deficiency: characterization of reproductive hormones over the menstrual cycle. *J Clin Endocrinol Metab*. 1989;69:804–812.
  46. Anttila L, Koskinen P, Irjala K, Kaihola HL. Reference intervals for serum sex steroids and gonadotropins in regularly menstruating women. *Acta Obstet Gynecol Scand*. 1991;70:475–481.
  47. Reginster JY, Sarlet N, Deroisy R, Albert A, Gaspard U, Franchimont P. Minimal levels of serum estradiol prevent postmenopausal bone loss. *Calcif Tissue Int*. 1992;51:340–343.
  48. Rothman MS, Carlson NE, Xu M, et al. Reexamination of testosterone, dihydrotestosterone, estradiol and estrone levels across the menstrual cycle and in postmenopausal women measured by liquid chromatography-tandem mass spectrometry. *Steroids*. 2011;76:177–182.
  49. Massart C, Gibassier J, Laurent MC, Le Lannou DL. Analytical performance of a new two-step ADVIA Centaur estradiol immunoassay during ovarian stimulation. *Clin Chem Lab Med*. 2006;44:105–109.
  50. Hegab HM, Schindler AE. The prognostic value of serum inhibin, 17  $\beta$ -estradiol and progesterone in cases of hydatidiform mole. *Gynecol Endocrinol*. 2004;18:107–113.
  51. Doria A, Iaccarino L, Sarzi-Puttini P, et al. Estrogens in pregnancy and systemic lupus erythematosus. *Ann NY Acad Sci*. 2006;1069:247–256.
  52. Randolph JF Jr, Zheng H, Sowers MR, et al. Change in follicle-stimulating hormone and estradiol across the menopausal transition: effect of age at the final menstrual period. *J Clin Endocrinol Metab*. 2011;96:746–754.
  53. Lee JS, Ettinger B, Stanczyk FZ, et al. Comparison of methods to measure low serum estradiol levels in postmenopausal women. *J Clin Endocrinol Metab*. 2006;91:3791–3797.
  54. Santen RJ, Demers L, Ohorodnik S, et al. Superiority of gas chromatography/tandem mass spectrometry assay (GC/MS/MS) for es-



- tradiol for monitoring of aromatase inhibitor therapy. *Steroids*. 2007;72:666–671.
55. Hsing AW, Stanczyk FZ, Belanger A, et al. Reproducibility of serum sex steroid assays in men by RIA and mass spectrometry. *Cancer Epidemiol Biomarkers Prev*. 2007;16:1004–1008.
  56. Khosla S, Melton LJ, Atkinson EJ, O'Fallon WM, Klee GG, Riggs BL. Relationship of serum sex steroid levels and bone turnover markers with bone mineral density in men and women: a key role for bioavailable estrogen. *J Clin Endocrinol Metab*. 1998;83:2266–2274.
  57. Khosla S, Amin S, Singh R, Atkinson E, Melton LJ III, Riggs B. Comparison of sex steroid measurements in men by immunoassay versus mass spectrometry and relationships with cortical and trabecular volumetric bone mineral density. *Osteoporos Int*. 2008;19:1465–1471.
  58. Auyeung TW, Lee JS, Kwok T, et al. Testosterone but not estradiol level is positively related to muscle strength and physical performance independent of muscle mass: a cross-sectional study in 1489 older men. *Eur J Endocrinol*. 2011;164:811–817.
  59. Vandendput L, Mellstrom D, Karlsson MK, et al. Serum estradiol is associated with lean mass in elderly Swedish men. *Eur J Endocrinol*. 2010;162:737–745.
  60. Trabado S, Maione L, Salenave S, et al. Estradiol levels in men with congenital hypogonadotropic hypogonadism and the effects of different modalities of hormonal treatment. *Fertil Steril*. 2011;95:2324–2329.e1–e3
  61. Kawashima H, Nakatani T. Involvement of estrogen receptors in prostatic diseases. *Int J Urol*. 2012;19:512–522.
  62. Agirbasli M, Agaoglu NB, Orak N, et al. Sex hormones, insulin resistance and high-density lipoprotein cholesterol levels in children. *Horm Res Paediatr*. 2010;73:166–174.
  63. Bernstein L, Lipworth L, Ross RK, Trichopoulos D. Correlation of estrogen levels between successive pregnancies. *Am J Epidemiol*. 1995;142:625–628.
  64. Pinheiro SP, Holmes MD, Pollak MN, Barbieri RL, Hankinson SE. Racial differences in premenopausal endogenous hormones. *Cancer Epidemiol Biomarkers Prev*. 2005;14:2147–2153.
  65. Vaidya D, Dobs A, Gapstur SM, et al. The association of endogenous sex hormones with lipoprotein subfraction profile in the Multi-Ethnic Study of Atherosclerosis. *Metabolism*. 2008;57:782–790.
  66. Yager JD, Davidson NE. Estrogen carcinogenesis in breast cancer. *N Engl J Med*. 2006;354:270–282.
  67. Key T, Appleby P, Barnes I, Reeves G; Endogenous Hormones and Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. *J Natl Cancer Inst*. 2002;94:606–616.
  68. Kaaks R, Rinaldi S, Key TJ, et al. Postmenopausal serum androgens, oestrogens and breast cancer risk: the European prospective investigation into cancer and nutrition. *Endocr Relat Cancer*. 2005;12:1071–1082.
  69. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. *J Natl Cancer Inst*. 2004;96:1856–1865.
  70. Key TJ. Endogenous oestrogens and breast cancer risk in premenopausal and postmenopausal women. *Steroids*. 2011;76:812–815.
  71. Onland-Moret NC, Kaaks R, van Noord PA, et al. Urinary endogenous sex hormone levels and the risk of postmenopausal breast cancer. *Br J Cancer*. 2003;88:1394–1399.
  72. Dorgan JF, Baer DJ, Albert PS, et al. Serum hormones and the alcohol-breast cancer association in postmenopausal women. *J Natl Cancer Inst*. 2001;93:710–715.
  73. Hamajima N, Hirose K, Tajima K, et al; Collaborative Group on Hormonal Factors in Breast Cancer. Alcohol, tobacco and breast cancer—collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. *Br J Cancer*. 2002;87:1234–1245.
  74. Beckmann L, Husing A, Setiawan VW, et al. Comprehensive analysis of hormone and genetic variation in 36 genes related to steroid hormone metabolism in pre- and postmenopausal women from the Breast and Prostate Cancer Cohort Consortium (BPC3). *J Clin Endocrinol Metab*. 2011;96:E360–E367.
  75. Allen NE, Key TJ, Dossus L, et al. Endogenous sex hormones and endometrial cancer risk in women in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Endocr Relat Cancer*. 2008;15:485–497.
  76. Lukanova A, Lundin E, Micheli A, et al. Circulating levels of sex steroid hormones and risk of endometrial cancer in postmenopausal women. *Int J Cancer*. 2004;108:425–432.
  77. Lee JS, Yaffe K, Lui LY, et al; Study of Osteoporotic Fractures Group. Prospective study of endogenous circulating estradiol and risk of stroke in older women. *Arch Neurol*. 2010;67:195–201.
  78. Chen Y, Zeleniuch-Jacquotte A, Arslan AA, et al. Endogenous hormones and coronary heart disease in postmenopausal women. *Atherosclerosis*. 2011;216:414–419.
  79. Laughlin GA, Kritz-Silverstein D, Barrett-Connor E. Endogenous oestrogens predict 4-year decline in verbal fluency in postmenopausal women: the Rancho Bernardo Study. *Clin Endocrinol (Oxf)*. 2010;72:99–106.
  80. Roddam AW, Appleby P, Neale R, et al. Association between endogenous plasma hormone concentrations and fracture risk in men and women: the EPIC-Oxford prospective cohort study. *J Bone Miner Metab*. 2009;27:485–493.
  81. Chatterton RT Jr, Parker NP, Habe-Evans M, Bryk M, Scholtens DM, Khan SA. Breast ductal lavage for assessment of breast cancer biomarkers. *Horm Cancer*. 2010;1:197–204.
  82. Hankinson SE, Manson JE, London SJ, Willett WC, Speizer FE. Laboratory reproducibility of endogenous hormone levels in postmenopausal women. *Cancer Epidemiol Biomarkers Prev*. 1994;3:51–56.
  83. Key T, Appleby P, Reeves G, et al. Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br J Cancer*. 2011;105:709–722.
  84. Roddam AW, Allen NE, Appleby P, Key TJ. Endogenous sex hormones and prostate cancer: a collaborative analysis of 18 prospective studies. *J Natl Cancer Inst*. 2008;100:170–183.
  85. Cummings SR, Browner WS, Bauer D, et al. Endogenous hormones and the risk of hip and vertebral fractures among older women. Study of Osteoporotic Fractures Research Group. *N Engl J Med*. 1998;339:733–738.
  86. McGuire WL. Estrogen receptors in human breast cancer. *J Clin Invest*. 1973;52:73–77.
  87. Janne O, Kontula K, Vihko R, Pystynen P, Auvinen O, Lauslahti K. Steroid binding properties of estradiol receptors in human breast cancer. *Med Biol*. 1975;53:214–223.